



Article

An Examination of the Failure Rate and Content Equivalency of Electronic Surrogates and the Implications for Print Equivalent Preservation

Ken Ladd
Associate Dean
University of Saskatchewan Library
University of Saskatchewan, Saskatoon, Canada
Email: ken.ladd@usask.ca

Received: 30 June 2010

Accepted: 23 Oct 2010

© 2010 Ladd. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 2.5 Canada (<http://creativecommons.org/licenses/by-nc-sa/2.5/ca/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

Abstract

Objective – This study sought to determine whether evidence indicates a need to preserve print equivalent journal collections. In addition, this research aimed to provide data on the failure rate of print equivalent materials for possible digitization to replace existing poor quality or defective electronic surrogates.

Methods – The project compared the content of randomly selected journal titles, volumes, and issues from seven electronic journal archives and their print equivalents held at the University of Saskatchewan Library. The archives were obtained from five separate vendors representing humanities, social sciences, science, technology, and medicine. Data were collected on the frequency and types of failure of electronic surrogates, supplemental content missing from electronic surrogates, and frequency and types of failure of print equivalent materials.

Results – Across all electronic journal archives the failure rate of electronic surrogates was 7.5% for all PDF documents and 11.5% for scholarly PDF documents. For individual electronic journal archives the failure rate ranged from 0.7% to 19.5% for all PDF documents and from 0.3% to 26.5% for scholarly PDF documents. Data is presented on the failure rate of individual electronic journal archives, types of failure, and missing supplemental content. An examination of print equivalent titles found 1.7% of print scholarly articles could not be used or were not optimal for digitization.

Conclusions – The study demonstrates the need for preserving print equivalent journal titles for at least the short (less than 5 years) to medium term (up to 10 years), while poorly digitized materials are identified, replaced, and digitally preserved. While electronic surrogates of image-rich scholarly papers are more likely to have quality issues, the study found some text-only PDF scholarly documents were illegible, indicating the need for caution against liberally applying this as a criterion for disposal of print equivalent titles. There is significant supplemental content absent from electronic surrogates which indicates a need for further discussion of the necessity for such information or for incorporating it into the digitization process to ensure a complete record of the print equivalent journals for future use. The failure rate of print equivalent titles for possible digitization provides additional data for discussions related to the determination of optimal overlap. It also suggests that the number of copies required for a full set of preserved journals over a specified time horizon may be greater than anticipated, unless page level validation is performed.

Introduction

Within the past decade there has been a massive migration of libraries moving from print to electronic resources. At the same time, academic research libraries have continued to face space issues related to their primary stacks and storage facilities due to simple growth in their collections. Other changes in the teaching and learning environment have resulted in an increased demand for social learning space within libraries. Coupled with the development of programs such as JSTOR, Portico, CLOCKSS, and publishers' electronic backfiles of serials, libraries have explored the storage or disposal of print journals when electronic surrogates exist.

Preservation is recognized as a fundamental role and responsibility of research libraries (Association of Research Libraries, 2007). The ultimate goal of these preservation activities is to ensure that the information contained within library collections is not lost for future access. However, with the shift toward acquiring electronic resources, there are questions about which titles need to be preserved or retained. Do both the print materials (print equivalents) and their electronic surrogates need to be preserved? What evidence exists to support either the disposal of print equivalents or their preservation? Are there certain indicators of whether both need to be preserved?

In 2006 the University of Saskatchewan Library, in collaboration with other units on campus, initiated a project to transform two floors of the Murray Library (primarily a humanities and social science library) to create an enhanced Learning Commons, requiring the removal of approximately five kilometres of shelving and associated contents.

With tight deadlines to remove materials, the library decided to discard print journals for which it had electronic surrogates. The Library determined the strategy also needed to incorporate preservation as a principle. The strategy was multi-faceted, with JSTOR titles being flagged for possible disposal. When other print equivalent titles' preservation was considered less reliable, those titles were flagged for an on-campus temporary storage facility. As the project was rolled out, there was considerable debate between library and other faculty regarding the storage or disposal of print equivalent titles. The spectrum of opinions ranged from both extremes – some felt that any print equivalent title could be discarded, while others felt no print titles should be discarded.

Literature Review

A variety of papers have examined either directly or peripherally the differences that exist between electronic surrogates and their print equivalents (Bracke & Martin, 2005; Campbell, 2003; Chen, 2005; Chrzastowski

2003; Erdman, 2006; Henebry, Safley, & George, 2002; Joseph, 2006; Kalyan, 2002; Keller, 2005; Martellini, 2000; Shadle, 2004; Sprague & Chambers, 2000). These studies were often performed to determine if libraries were able to cancel or withdraw print equivalents from their collections. However, these studies have tended to focus on a specific discipline, specific content issue, specific vendor, or electronic journal databases and aggregators, not the electronic journal itself.

Studies that have been more multidisciplinary in nature include Sprague and Chambers (2000), Kalyan (2002), and Chen (2005). These studies compared electronic surrogates from multidisciplinary full-text databases with their print equivalents. Studies by Keller (2005) and Henebry, Safely and George (2002) examined electronic journals from a variety of publishers and disciplines. Many of the studies, however, have focused on the sciences:

- Chemistry - Chrzastowski (2003)
- Earth and planetary sciences - Joseph (2006)
- Geology - Erdman (2006)
- Physics - Martellini 2000
- Science and engineering - Bracke and Martin (2005)
- Science, technology, and medicine – Campbell (2003)

Studies have found that there are often quality issues associated with digitized images and figures (Bracke & Martin, 2005; Chen, 2005; Erdman, 2006; Henebry, Safely, & George, 2002; Joseph, 2006; Keller, 2005; Sprague & Chambers 2000), while others have not (Campbell 2003; Chrzastowski 2003). Few studies have indicated a quality issue with text in electronic surrogates. Sprague and Chambers (2000) noted that formulas and other mathematical expressions were often unclear. Keller (2005) noted that some pages of electronic surrogates were not readable.

In addition to the issue of quality, some studies have examined or noted the issue of missing content:

- Missing figures, tables, or graphics – Chen (2005); Sprague and Chambers (2000)
- Missing pages or articles – Bracke and Martin (2005); Henebry, Safely, and George (2002); Keller (2005)
- Missing issues – Bracke and Martin (2005); Joseph (2006); Keller (2005)
- Missing volumes – Joseph (2006); Keller (2005)

Two studies did not find an issue with missing content (Martellini, 2000; Chrzastowski, 2003).

Keller (2005) and Shadle (2004) identified inconsistencies in the journal titles as presented by publishers on their websites, which might be confusing to users. Some publishers are very good at noting title changes, while for others only the most current title is displayed.

There have been several studies that have focused either on full text journal databases and aggregators (Chen, 2005; Kalyan, 2002; Sprague & Chambers 2000) or on a specific vendor – Elsevier (Bracke & Martin, 2005; Erdman, 2006; Joseph, 2006).

With the placement of numerous back runs of journals into its temporary storage facility, and the need to store other back runs permanently, the University of Saskatchewan Library was interested in exploring opportunities for a collaborative approach for the preservation of print journals. Preservation of print journals is complicated, because many factors need to be considered in determining the redundancy required to ensure the existence of a complete run of a journal. Schonfeld and Housewright (2009) discussed these factors, including the work commissioned by Ithaka S+R (Yano, Shen, & Chan, 2008), which noted the redundancy required is dependent on a number of risk factors, including defects in the print materials and loss.

The present study was initiated in the spring of 2009 to systematically compare print journals and their electronic surrogates from a variety of vendors across all disciplines. The

study expands on the existing literature by identifying and quantifying discrepancies between print journals and their electronic surrogates. It also quantifies damage or other irregularities in print journals that limit their use for digitization purposes.

Aim

The primary purpose of this study was to examine whether there is evidence to support the null hypothesis that there is a need to preserve print equivalent serials, at least for the short to medium term. To examine this issue the following questions were asked:

- What types and frequency of failure occurred for electronic surrogates?
- Were there differences in failure rates between electronic surrogate archives?
- In addition to failures, what content differences existed? For example, tables of contents, indices, advertising, etc.
- What other issues affected access to electronic surrogate content?

A secondary purpose was to examine whether the suggestion made by Yano, Shen, and Chan (2008) on the defect rate of print resources for digitization purposes was supported by evidence from the collection at the University of Saskatchewan library. To examine this issue, data were collected on the types of failures that occurred with print equivalents and the failure frequency.

Methods

At the time the study was initiated, the University of Saskatchewan Library had acquired access to 28 archive collections of electronic journal backfiles from a number of vendors. From these collections, seven were chosen to provide a breadth of subject coverage (humanities, social sciences, science, technology, and medicine) from a variety of vendors. Appendix A provides a listing of the

collections and which of their journals were included in this study.

From each collection, five titles held at least partially in print by the University of Saskatchewan library were randomly selected. From each of these titles, a specific number of volumes was randomly selected. The number of random volumes was pre-determined based on the library's holdings and ranged from one to three volumes, but was usually three (60%). One exception occurred when six volumes for a title were selected because each volume had only one issue. Within the volumes, a random number of specific issues was selected, based on the number of issues in the volume, as indicated in the electronic surrogate. Where there were three or fewer issues in a volume, one issue was selected; with four to twelve issues in a volume, two issues were selected. Rare exceptions occurred when combined issues for a volume were randomly selected.

For each randomly selected journal issue, the electronic surrogate and the print equivalent were examined. Study data for each collection and journal title included several aspects of failure:

- The frequency that an electronic surrogate failed at the article (or PDF document) level was defined as any time the print equivalent needed to be consulted to access all the information in the item. While there could be multiple failures within a PDF document, together they would be counted as a single failure for the journal title and collection.
- The failure rate for each journal was determined for all PDF documents as well as for all "scholarly" content. For this study, scholarly content included research papers, case studies, review articles, short communications, technical notes, and errata. In addition to the "scholarly content," "all content" included book reviews,

announcements, letters to the editor, meeting programs, and obituaries. A distinction was made between the two content types, because users seek the scholarly content of scholarly journals most often. Thus, a failure in a scholarly PDF document was more likely to result in the print equivalent being required.

- The type of failure was observed for electronic surrogates, and it was recorded for only the initial failure in a PDF document. Eight categories were used to describe the observed failures. Five of the categories related to the quality of the scan or of the digitization. The other three categories related to missing, incorrect, or additional content.
- Frequency and types of supplemental content missing from the electronic surrogate were noted. Supplemental content was defined as anything beyond the article level. Examples could be advertising, tables of contents, instructions for authors, and so on. On a single page more than one type of supplemental content

could occur, and each was recorded.

- The research team noted the frequency that print equivalent articles failed as a possible source for digitization. Failure was defined as any occurrence of missing or damaged pages, such as markings or tears.
- The study also identified several types of failure for print equivalent articles.

Each digital surrogate within an issue was examined for content equivalency and the legibility of text, graphs, figures, and images. The digital surrogate was first examined for quality issues and any perceived problems were compared to the print equivalent to determine if they were an artifact of the digitization process. A second examination evaluated the print equivalent for quality issues, while at the same time noting irregularities with content equivalency.

Results

The study involved an examination of seven archival electronic journal collections covering the humanities, social sciences, science, technology, and medicine. Table 1 indicates the number of titles in each collection and the number of titles sampled.

Table 1
Electronic Archival Collections Examined

Collection	Titles		Volumes			PDF
	Archive	Sample	Archive	Held	Sample	Compared
Elsevier - Medicine and Dentistry	44	5	128	104	14	344
Elsevier - Social Science	26	5	72	54	12	455
JSTOR Arts and Science I	175	5	311	301	14	701
Oxford University Press Digital Archive	50	5	355	339	18	553
Springer Link Historical Archives Mathematics	34	5	32	26	8	53
Wiley - Humanities and Social Sciences	69	5	169	118	13	247
Wiley - Science, Technology and Medicine	79	5	105	97	12	280
TOTAL	477	35	1,172	1,039	91	2,633

For the titles sampled, the table notes the number of volumes within each electronic archive held in the University of Saskatchewan library that were sampled. Finally, the table indicates the number of electronic surrogate PDF documents that were compared with their print equivalents.

The initial part of the study examined the quality of the sampled electronic surrogates by determining failure rate – how frequently print equivalent materials had to be consulted in order to access all the information contained in the item. Each vendor was found to follow digitization practices for print equivalent journals that impacted the failure rate for all PDF documents. For some vendors the number of PDF documents was higher because individual PDF files were created for each book review, table of contents, obituary, and announcements, in addition to the scholarly articles, case studies, review articles, errata, and short communications. For example, whether a vendor used one PDF document for ten book reviews or presented them as ten PDF documents impacted the calculation of the failure rate.

The study found a wide variance of failure rates between backfile collections. Figure 1 shows the percentage failure rate of scholarly PDF documents and of all PDF documents for each of the seven archival collections and across all collections.

Five of the seven collections had higher failure rates for scholarly PDF documents when compared to all PDF documents. There was a wide variance in failure rate between collections, with the science, technology, and medicine collections usually having the highest failure rates. For scholarly PDF documents, JSTOR's failure rate was at least an order of magnitude lower than any other collection. Three of the collections had higher failure rates for scholarly PDF documents than the average rate of 11.5% observed for all seven collections. While JSTOR's mission differs from that of commercial vendors, the results demonstrate the quality that can be achieved with journal digitization initiatives, which logically can be ascribed to excellent quality control practices.

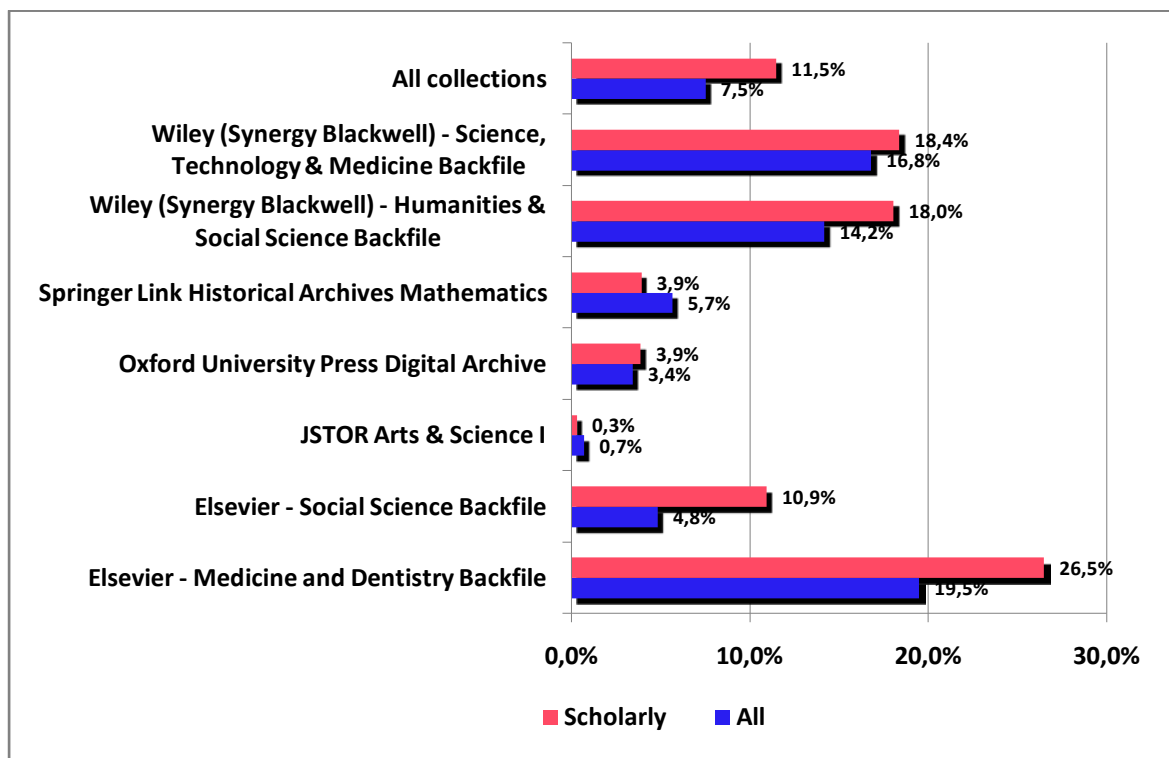


Fig. 1. Failure rate for electronic surrogates (scholarly and all PDF documents) when compared to print equivalents.

The study examined types of observed failures (Fig. 2).

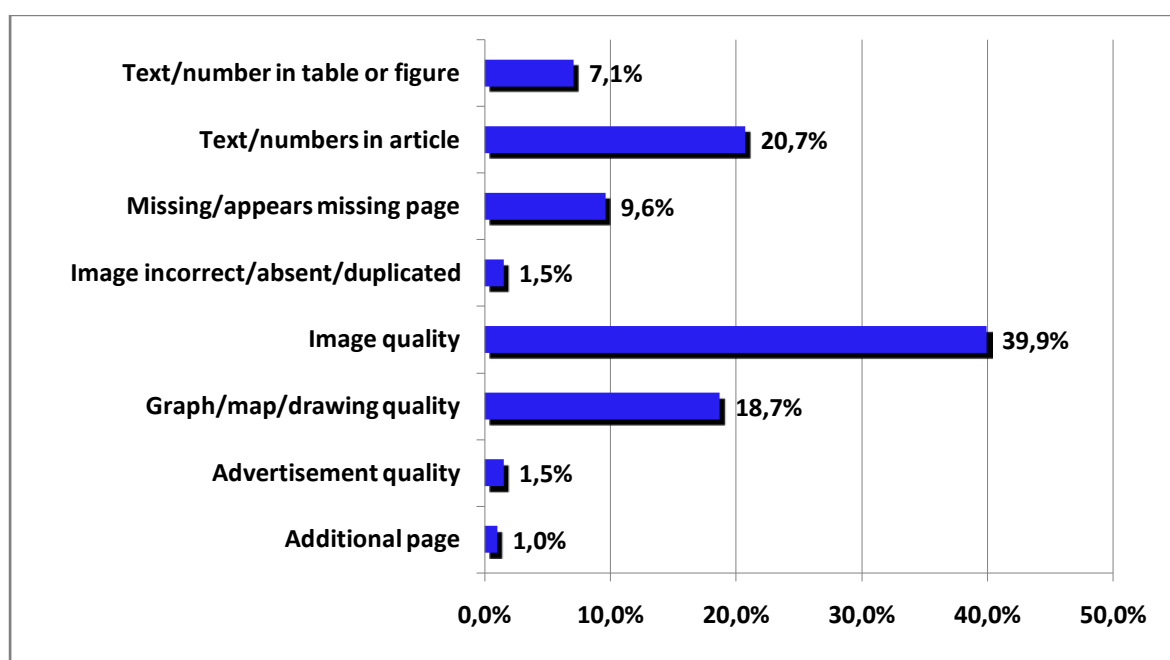


Fig. 2. Types of failures observed first within an electronic surrogate PDF file as a percentage of all observed failures.

The predominant type of failure (39.9%) was related to the quality of an image, such as an x-ray, photomicrograph, chromatograph, or scintograph. Other quality-related issues were observed for text and numbers: 20.7% of these failures occurred in the body of a paper, and 7.1% occurred in a table or figure. The quality of graphs, maps, and drawings was an issue in 18.7% of failures resulting from difficulty distinguishing different shading or fill for a bar graph, symbols on a graph, or lines on a graph, drawing, or map. Missing pages accounted for nearly 10% of the failures. For two cases there appeared to be missing pages, but closer inspection determined that one page appeared as a miniature image within the PDF. By clicking on the image and increasing magnification to 2400% or higher, it was possible to read the content. While only the initial failure was recorded, there were 15 PDF documents that had a second type of failure, or 7.6% of all failed PDF documents.

One observation of importance not included in the calculations for failure rates or types of failures was the complete absence of electronic surrogates for two issues from one volume of

a journal, or 1.1% of all volumes sampled for the study. The *International Journal of Nuclear Medicine and Biology* is part of the Elsevier Medicine and Dentistry collection. The random selection of volumes and issues was based on the electronic surrogate journal issues listed by the vendor. In this case, the vendor's site indicated four issues for volume 12. After a comparison with the print equivalent, six issues were identified as having been published in this volume. The missing issues contained one editorial, twenty-two papers, one technical note, three letters to the editors, reports of eleven new patents, one book review, one announcement, and advertising.

The study compared the electronic surrogate PDF documents associated with a journal issue and their print equivalent journal issues to determine whether there was supplemental content not included as an electronic surrogate. The research team analyzed missing pages for all journal issues selected for this study, and results were graphed as a percentage of all journal issues examined (Fig. 3).

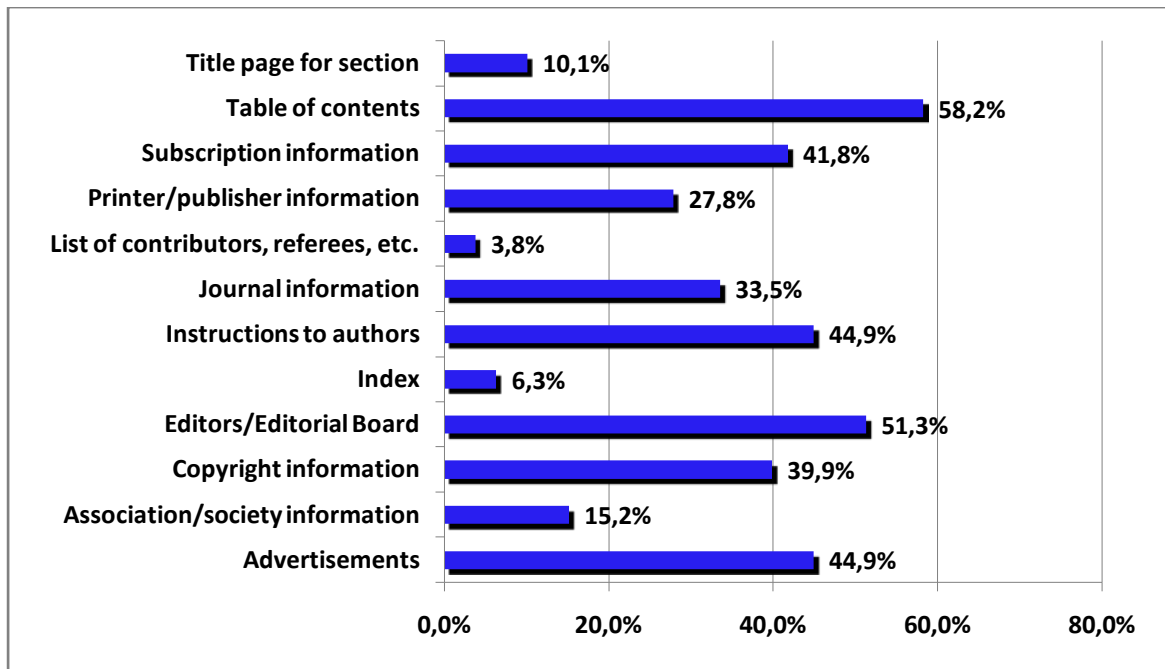


Fig. 3. Supplemental content found in print equivalent journal issues not present in electronic surrogates.

The study found eight different types of missing supplemental content in over 25% of the journal issues examined, with two types being above 50%.

One function of a print archive can be digitization at the article level to replace poor quality or defective electronic surrogates. Data collected on the failure of print equivalent materials are shown in Table 2.

Failure of print equivalent materials at the individual document level was quite low. For all print equivalent scholarly documents, the failure rate was 1.7% (26 of 1,552 items). For all print equivalent documents examined

(scholarly and other), the failure rate was 1.3% (34 of 2,633 items).

Over half of the failures (n=14) for scholarly print equivalent documents were the result of articles being marked up by users with pens, pencils, or highlighters. There were 6 damaged issues, primarily due to pin binding, and two papers that had been removed intentionally. (One paper had been completely removed, resulting in a page missing from the preceding paper in the same issue.) Two papers could not be read as bound, but could perhaps have been used for digitization purposes if unbound. Finally, there were two papers that were illegible due to poor print resolution.

Table 2
Occurrences of Types of Print Failures

Type of failure	Scholarly	Other
Tight or close binding, not legible	2	0
Print faded or otherwise not legible	2	0
Damaged page	6	2
Missing page	2	3
Page markings	14	3
TOTAL	26	8

At the issue level, the highest frequency of print failure was associated with the removal of covers or pages for binding. A total of 19% (30 of 158) physical issues had this material removed.

In addition to the problems associated with electronic surrogates noted above, there were a variety of other problems observed. For some vendors the only journal title listed for an electronic surrogate was the current title. If there had been one or more title changes, the previous titles were not listed. Many vendors had pagination errors associated with a particular issue or with individual PDF documents. A total of 57 pagination errors were noted, with 54.4% of these errors being associated with scholarly PDF documents.

Discussion

Failure of Electronic Surrogates

One of the challenges associated with comparing electronic surrogates with their print equivalents is the subjectivity involved. At what point does an electronic surrogate fail by requiring the user to access the print equivalent? For this study a somewhat liberal definition was used – if the text was not clearly legible but the word(s) in question could be deduced from the context provided by other legible text, it was not considered a failure. Thus, the failure rate of 11.5% for all scholarly PDF documents can be considered conservative.

A significant observation of the study was the absence of corresponding electronic surrogates for two issues of one Elsevier journal title and volume. Because the study design used the electronic surrogates as the basis for randomly selecting journal issues, the two missing journal issues were not included for possible random selection and therefore not incorporated into the calculation of failure rate of electronic surrogates. If either of these issues had been included in the study, the observed failure rate for scholarly PDF documents would have increased by approximately 0.5% to 12%.

In examining the different failure rates by collection, it was evident that for collections containing image-intensive papers, there was a corresponding increase in failure rates. This interpretation was supported by the examination of types of failures, where approximately 40% of all initial failures were due to poor image quality. Including failures associated with poor quality graphs, maps, and drawings increased the failure rate to almost 60% for poor quality images and figures. This result supports the use of image intensiveness of publications as a criterion for preservation, as noted in Ithaka's recent paper on what to withdraw (Schonfeld & Housewright, 2009).

At the other end of the spectrum are text-only publications and the possible use of this criterion for materials that could be potentially withdrawn with minimal risk. While the study confirmed that there are certainly fewer problems associated with text-only publications, they are not without significant problems. Of the initial problems, approximately 20% of the failures were associated with text. For one particular title, *Papers in Regional Science* (current title), the failures were frequently associated with mathematical formulas where there were super- or sub-scripts. However, for this title the overall quality of the electronic surrogate was poor, making text difficult to read in general. If the text associated with tables and figures had been included, the overall failure rate related to text would have increased to almost 28%.

The study found that many of the failures appeared to be with earlier volumes for titles. This might imply that the digital surrogates were created with technology that produced lower resolution or quality of electronic surrogates. Re-digitization of these materials would likely eliminate many of the failures, especially for text-only titles.

While the quality of the electronic surrogate is something that can be addressed by the re-digitization of the print equivalents, there also is an overall quality control issue. Almost 10%

of all initial failures were associated with pages missing from the electronic surrogate. An additional 1.5% of the failures were associated with incorrect digital content. Each of these issues is related to quality control applied following the digitization of the materials. There is an obvious cost associated with applying high quality control practices, but even a minimal inspection should catch many of the observed failures, such as incorrect pages scanned or missing pages. It was apparent that JSTOR, at least for the titles examined, has incorporated quality control practices to ensure high-resolution electronic objects and almost no failures. Thus, the issue appears to be more about quality control and the use of high-resolution digitization technology than whether a title is text-only or image-intensive.

Supplemental Content Missing from Electronic Surrogates

When comparing print equivalents and electronic surrogates, the issue of content equivalency emerges. Steve McKinzie (2005) argued that back runs of print equivalent journals should be kept, as electronic surrogates may not include advertisements, conference announcements, and other material. This study examined the scope of this issue.

Any content, beyond scholarly work, book reviews, letters to the editor, editorials, and so on was noted. This was not intended to be a comment on the value of this supplemental content but rather an inventory of the type of material being excluded. It should be noted, however, that in some cases vendors do provide this content, often as "front matter" or "back matter," so in those cases others have made some value judgments.

For 58.2% of the issues examined, the table of contents was not provided as an electronic surrogate. As vendors often provide a table of contents in the form of a list of electronic surrogates for a particular issue, this may be less of an issue. Other types of supplemental content that may have less impact when

absent are indexes and title pages for sections in an issue.

There was a variety of other supplemental content not included that historians, sociologists, librarians, and others might wish to consult. This included information related to editors and editorial boards, which was missing from over 50% of the electronic journal issues examined for the study. This information would be needed to identify editorial changes that have occurred over time. Advertisements were absent 44.9% of the time and would likely be of interest for sociological research. Information about the journals such as aims and scope, objectives, editorial policy, indexing resources, and availability of back and special issues was absent more than one-third of the time. Missing information about the association or society publishing the journal included directories of officers, membership lists, and association histories. Other missing content of possible interest included instructions to authors and lists of contributors, reviewers, and referees.

The impact of supplemental content being absent is dependent on the perceived value of the information and the need to access it. It could be argued that the information might be used for research or general information purposes. If we continue to move towards reliance on electronic surrogates and the disposal of print, the potential impact of its absence increases. Thus, it would be beneficial to include this information in the electronic surrogate collections to ensure that options are not limited in this area.

Failure Rate for Print Equivalent Journals

Two variables of interest to discussions of optimal overlap are the defect rate and the loss rate of print equivalent journals that could be used for digitization (Yano, Shen, & Chan, 2008). Defects could include damage (e.g., pages removed intentionally, marked pages, or torn pages). The loss of content, whether through defect or loss, impacts on access and the ability to digitize.

In determining the suitability of the print equivalents examined in this study for digitization purposes, it was found that overall the worst-case scenario was a failure rate of 1.7% for scholarly print equivalent documents and 1.3% for all print equivalent documents. The most frequently observed damage to content was from users marking the item for their own use, which occurred in more than 50% of the instances, or 0.9% of the print equivalent documents. In most instances the markings did not obscure the content, indicating digitization would be possible but not optimal.

Damage to pages was the second most frequent observation for print failure. At the University of Saskatchewan this was primarily from the past practice of occasionally using pin binding. Holes drilled for pin binding at times went through text, rendering it unusable for digitization purposes.

At the outset of the study, there was some initial concern about the level of damage that might be observed from users intentionally removing content. Surprisingly, only one journal issue was found to have materials intentionally removed, which represented only 0.1% of the print equivalent scholarly content. While the individual was targeting only one scholarly paper, its removal resulted in four print equivalent documents being affected – one page from a second article, the targeted article, and two non-scholarly documents. The observed level of intentional removal of content may be due in part to the journals that were randomly selected, as anecdotal evidence indicates that certain collections that do not circulate at the University of Saskatchewan Library (such as Nursing) or those that contain art-related images are more susceptible to intentional damage.

In the two instances where the print equivalent materials were not legible, the problem was due to poor printing processes used for the original documents. The electronic surrogate for these items was also illegible, resulting in the loss of some content.

As a result, re-digitization would not be an option. For the two cases where pin binding made the materials illegible, removing the binding might have resolved the issues for digitization purposes.

Yano, Shen, and Chan (2008) noted that for 23 of 25 JSTOR journals being prepared for scanning, there was a defect rate of one per 10,000 to 100,000 pages. The 2 journals with a higher defect rate were a nursing journal and a medical journal. The authors speculated that the defects might have been due to higher usage. They noted that these statistics might not represent journals in general, because JSTOR had sought copies that were relatively clean. They observed that a significant portion of these materials were obtained from major research university libraries and suggest that “off-the-shelf” journals will generally be of very good condition. The results of this study cannot be directly compared; however, the defect rate appears to be higher here with 1.7% of scholarly print equivalent documents failing. These results suggest that the condition of “off-the-shelf” journals will generally be good, but this risk factor is higher in the current study.

Other Observed Issues

In comparing print equivalents with the listing of electronic surrogates provided by vendors, there were frequent errors associated with the pagination listed for the surrogates. This could be potentially confusing to users trying to locate a specific paper. In most cases it would only be an obstacle, as an examination of article titles and authors should result in accessing the desired paper. It also reflects an overall quality control problem that may be a flag for other issues.

Perhaps a more confusing problem, particularly for individuals who are not familiar with the journal in question, is the practice of some vendors of not noting title changes. Both JSTOR and Elsevier were particularly good at tracing title changes, while Wiley and Oxford University Press were less so. This issue of inaccurate journal titles

has been noted before (Shadle, 2004; Keller, 2005), but has not been resolved in the intervening years. Librarians and users must continue to work with publishers and vendors to ensure they are aware of the importance of recording title evolution to ensure users can easily locate the resources they require, and to enable libraries to confidently identify their electronic surrogate collections.

Conclusion

This study was initiated to determine whether evidence supported the preservation of print equivalent journal collections. Evidence was sought by determining how frequently electronic surrogates failed to provide access to all content within an individual PDF document. Of particular interest was the failure rate associated with the scholarly content of journals. Recording the types of failures provided evidence associated with the use of different criteria for the preservation or withdrawal of print equivalent journals.

In addition, the study examined the frequency that print equivalent materials failed to be eligible for digitization to replace poor quality electronic surrogates. Recording the frequency and types of failures provided evidence associated with calculating optimal overlap for archiving of print journals.

The study clearly demonstrates there is a need for preserving print equivalent journal titles for at least the short to medium term. While the electronic surrogates of image-rich scholarly papers are more likely to have quality issues, the study found some text-only scholarly papers were illegible, indicating caution for liberally using this criterion for disposal of print equivalent titles. This is further supported by ample evidence of quality control related issues, such as duplicate pages, missing pages, missing issues, additional pages, and poor quality scans. Re-digitization with high-resolution scanning technology and good quality control practices would eliminate many of the observed failures.

Retaining print equivalent journals for the short to medium term will place additional pressure on libraries already facing space issues related to expanding collections and the demands for user-related space. This pressure will be best met by collaborative approaches to retaining materials at the regional or national level.

The absence of supplemental content in many cases indicates the need for further discussion of the necessity for such information or incorporating it into the digitization process to ensure a complete record of the print equivalent journals for future use.

The failure rate of print equivalent titles suitable for digitization provides additional evidence of defect rates that applies to work by Yano, Shen, and Chan (2008) on optimal overlap for print preservation models. This study's results indicate the risk factor was greater than that noted by Yano and her colleagues for titles being prepared for scanning by JSTOR. Thus, the number of copies required for a full set of preserved journals over a specified time horizon may be greater than anticipated, unless page level validation is performed.

While this study demonstrates a variety of deficiencies related to electronic surrogates of print equivalent journals, a future study of the impact of these deficiencies on libraries and their users would be useful. Determining the impact will indicate the risks associated with not addressing these deficiencies and will assist decision-making related to digitization, preservation, and retention of print equivalent volumes. In addition, a study that quantifies the issue of web-based title inconsistencies would be helpful. Although several studies, including this paper, have observed and commented on this issue, it has not been quantified. Such a study should shed light on the extent of the problem and explore current practices of specific publishers regarding tracing title histories and best practices.

References

- Association of Research Libraries. (2007). *Research libraries' enduring responsibility for preservation*. Retrieved 25 Nov. 2010 from http://www.arl.org/bm~doc/preservation_responsibility_24july07.pdf
- Bracke, M. S., & Martin, J. (2005). Developing criteria for the withdrawal of print content available online. *Collection Building*, 24(2), 61-64.
- Campbell, S. (2003). Print to electronic journal conversion: Criteria for maintaining duplicate print journals. *Felicitier*, 49(6), 295-297.
- Chen, X. (2005). Figures and tables omitted from online periodical articles: A comparison of vendors and information missing from full-text databases. *Internet Reference Services Quarterly*, 10(2), 75-88.
- Chrzastowski, T. E. (2003). Making the transition from print to electronic serial collections: A new model for academic chemistry libraries? *Journal of the American Society for Information Science and Technology*, 54(12), 1141-1148.
- Erdman, J. M. (2006). Image quality in electronic journals: A case study of Elsevier geology titles. *Library Collections, Acquisitions, and Technical Services*, 30(3-4), 169-178.
- Henebry, C., Safley, E., & George, S. E. (2002). Before you cancel the paper, beware: All electronic journals in 2001 are NOT created equal. *The Serials Librarian*, 42(3-4), 267-273.
- Joseph, L. E. (2006). Image and figure quality: A study of Elsevier's earth and planetary sciences electronic journal back file package. *Library Collections, Acquisitions, and Technical Services*, 30(3-4), 162-168.
- Kalyan, S. (2002). Non-renewal of print journal subscriptions that duplicate titles in selected electronic databases: A case study. *Library Collections, Acquisitions, and Technical Services*, 26(4), 409-421.
- Keller, A. (2005). The race to digitize: Are we forfeiting quality? *Serials*, 18(3), 211-217.
- Martellini, E. (Oct. 2000). Physics journals and their electronic version: A comparison. *High Energy Physics Libraries Webzine*, (2), Retrieved 20 Nov. 2010 from <http://library.web.cern.ch/library/Webzine/2/papers/3/>
- McKinzie, S. (2005). Op ed: Troubling choices: Full-text access and the old hard copy back runs. *Against the Grain*, 17(1), 60-61.
- Schonfeld, R. C., & Housewright, R. (29 Sept 2009). *What to withdraw? Print collections management in the wake of digitization* ITHAKA. Retrieved 25 Nov. 2010 from <http://www.ithaka.org/ithaka-s-r/research/what-to-withdraw>
- Shadle, S. (2004). Electronic journal forum: Reflections on wrapping paper: Random thoughts on AACR2 and electronic serials. *Serials Review*, 30(1), 51-55.
- Sprague, N., & Chambers, M. B. (2000). Full text databases and the journal cancellation process: A case study. *Serials Review*, 26(3), 19-31.
- Yano, C. A., Shen, Z. J. M., & Chan, S. (2008). JSTOR seeks efficiency and security for print backups of online journals. Berkeley, CA: Department of Industrial Engineering and Operations Research, University of California.

Appendix A: Titles Compared in Each Collection

Elsevier – Medicine and Dentistry

- ✓ *American Journal of Orthodontics*
- ✓ *Biochemical Medicine and Metabolic Biology*
- ✓ *British Journal of Tuberculosis and Diseases of the Chest*
- ✓ *International Journal of Nuclear Medicine and Biology*
- ✓ *Prostaglandins, Leukotrienes, and Medicine*

Elsevier – Social Sciences

- ✓ *Government Publications Review*
- ✓ *Journal of Behavioral Economics*
- ✓ *Social Science & Medicine. Part B, Medical Anthropology*
- ✓ *Studies in Comparative Communism*
- ✓ *Transportation Research. Part A, General*

JSTOR Arts and Sciences 1

- ✓ *American Journal of Mathematics*
- ✓ *Journal of Health and Human Behavior*
- ✓ *Journal of the History of Ideas*
- ✓ *Reviews in American History*
- ✓ *Speculum*

Oxford University Press

- ✓ *Occupational Medicine*
- ✓ *Parliamentary Affairs*
- ✓ *Past & Present*
- ✓ *Rheumatology*
- ✓ *The Year's Work in Clinical and Cultural Theory*

Springer Mathematics

- ✓ *Computational Optimization and Applications*
- ✓ *Constraints*
- ✓ *Journal of cryptology*
- ✓ *Journal of nonlinear science*
- ✓ *K-Theory*

Wiley Interscience (Synergy Blackwell) – Humanities and Social Sciences

- ✓ *Papers in Regional Science*
- ✓ *Social Policy and Administration*
- ✓ *Journal of Philosophy of Education*
- ✓ *Psychology of Women Quarterly*
- ✓ *Review of Policy Research*

Wiley Interscience (Synergy Blackwell) – Science, Technology and Medicine

- ✓ *European Journal of Clinical Investigation*
- ✓ *International Journal of Experimental Pathology*
- ✓ *Journal of Human Nutrition and Dietetics*
- ✓ *Journal of Oral Pathology and Medicine*
- ✓ *Sedimentology*